

# Conversione digitale dei cataloghi

## Dalle schede all'Opac



**Marco Martini**  
**Biblioteca Centrale "G. Marconi"**

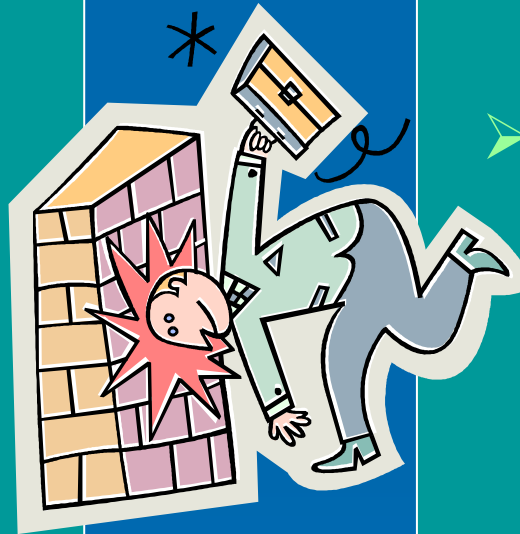
# Sommario

- Il problema conversione
- Catalogo
- Opac e Metaopac
- Tipologie di accesso agli Opac
- Conversione del catalogo cartaceo  
Tecniche e strategie
- Tecnologia di lettura ottica

➤ Catalogo cartaceo

➤ Circa 210.000  
schede  
bibliografiche

➤ Circa 85.000  
schede principali  
da riversare



➤ Opac BICE

➤ Circa 73.000  
record bibliografici  
immessi dal 1987

# Perché recuperare il pregresso

- Consentire o facilitare la consultazione del catalogo agli utenti
  - Interni, che non devono ripetere le ricerche su più sistemi di ricerca (manuali e non)
  - Esterni, che possono consultare a distanza
- Implementare gli standard catalografici
- Agevolare la pubblicazione del catalogo e dei suoi derivati su diversi tipi di supporto

# Catalogo

- Elenco ordinato dei documenti conservati in una biblioteca, descritti sulla base di norme coerenti e principi uniformi con lo scopo di mettere in comunicazione l'utente con gli stessi e con le informazioni che essi contengono.

# Tipi di cataloghi

- A ciascun documento corrisponde un'unica descrizione che può essere indicizzata e recuperata
  - secondo varie opzioni di ricerca (per autore, per soggetto, per classe)
  - Secondo sequenze diverse (alfabetiche o numeriche)

# Tipi di cataloghi II

- Sulla base dell'elemento scelto come indice si distingue tra
  - Catalogo per autore
  - Catalogo per soggetto
  - Catalogo per classe o sistematico
  - Catalogo dizionario

## Catalogo cartaceo



Separazione fisica tra i vari tipi di cataloghi. Ogni via di accesso ad uno stesso documento deve essere presentata insieme con la sua descrizione che sarà ripetuta tante volte per ciascuna delle vie di accesso.

## Catalogo automatizzato



Sistema informativo unitario. Presenza quindi di un unico strumento dotato di svariate categorie di accesso all'archivio che contiene le descrizioni dei documenti.



# OPAC

- On-line Public Access Catalogue
- E' l'insieme della base dati con il suo sistema di gestione, è quindi il catalogo con tutti i suoi record strutturati in campi definiti.
- Nasce dalla fusione del catalogo con la sua interfaccia di accesso.

# Funzione dell'OPAC

- Dare accesso a una base dati, cioè al catalogo che sta dietro
- Rendere l'informazione direttamente fruibile da parte dell'utenza
- Limitare la rigidità di un sistema gestionale strutturato in modo poco comprensibile
- Permettere di recuperare dati pertinenti alle query poste, restituendo degli output organizzati in form comprensibili
- Realizzare la funzione primaria di mediazione tra base dati e l'utente

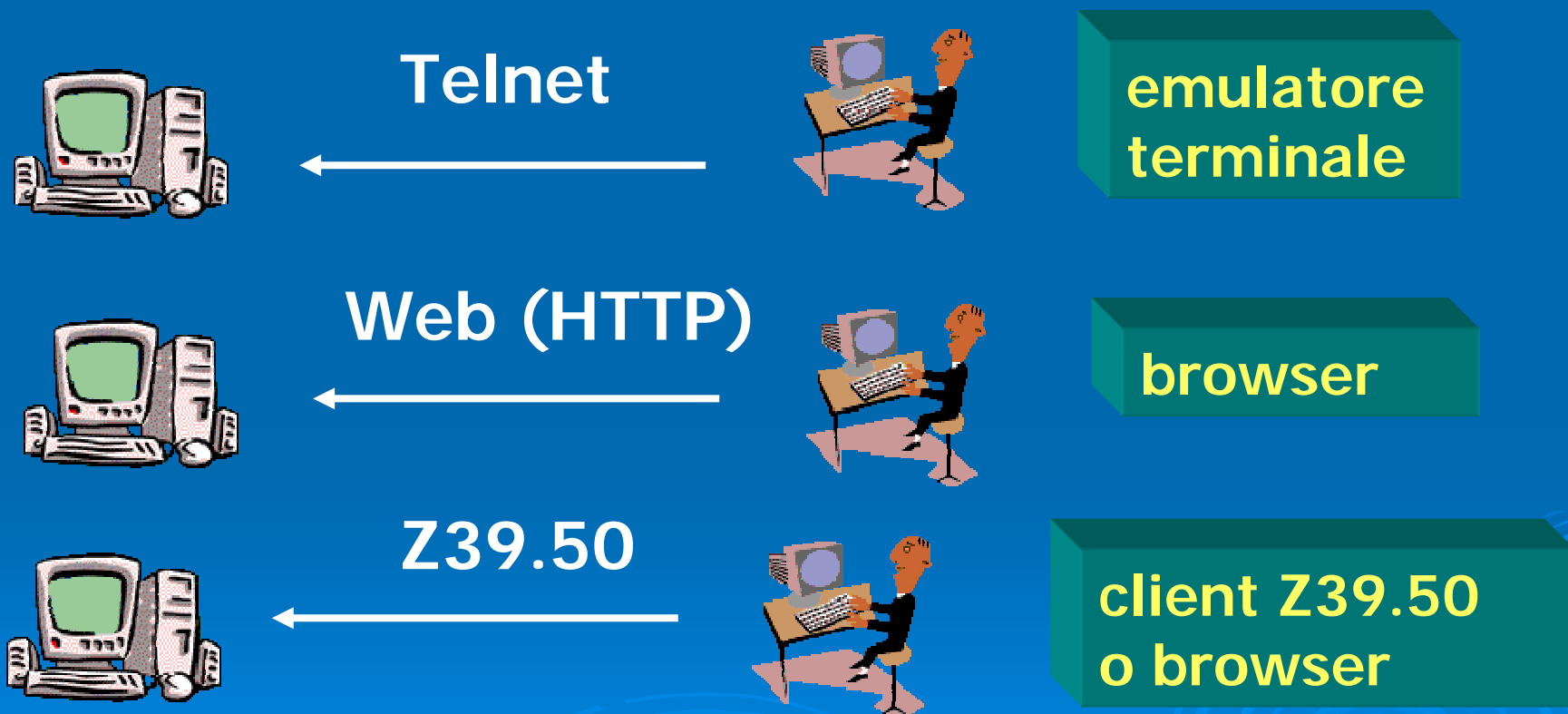
# Interfaccia dell'OPAC

- L'interfaccia di un OPAC è il prodotto attraverso il quale l'archivio elettronico (o database), residente sul server locale e contenente le registrazioni bibliografiche costantemente alimentate dall'attività catalografica dei bibliotecari, dialoga con l'utente. L'interfaccia è quindi lo strumento visivo (la maschera di interrogazione) attraverso il quale il database (*server*) si presenta all'utente (*client*).

# Interfaccia dell'OPAC II

- La tipologia di interfaccia determina la distinzione tra OPAC testuali e OPAC grafici. Entrambe sono accessibili tramite un pc connesso alla rete ma ciò che le distingue è il protocollo di connessione alla rete che utilizzano per la trasmissione dei dati.

# Modalità di accesso all'OPAC



# Modalità di accesso all'OPAC

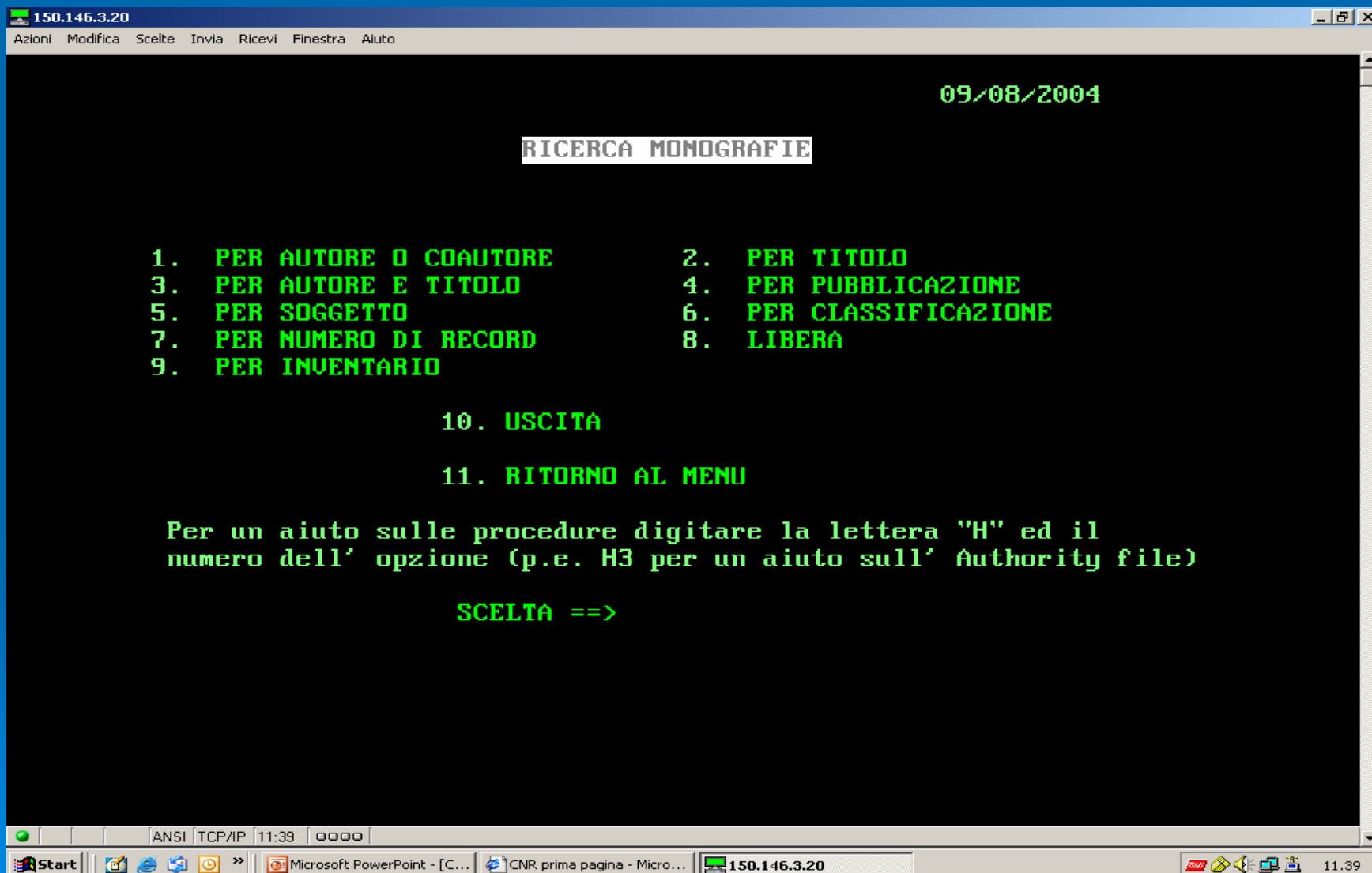
## Opac testuale - Telnet

- Chiamata ad un indirizzo fisico (IP Address)
- Protocollo Telnet VT100
- Interfaccia a carattere
- Modalità di ricerca strutturata
- Raffinamento per fasi successive

# Telnet

- Il comando Telnet consente in pratica di collegarsi ad un calcolatore remoto in emulazione di terminale (remote login), cioè come se il proprio personal computer fosse un terminale di quel calcolatore.

# Modalità testuale - Telnet





# Modalità di accesso all'OPAC

## Opac Web

- E' un catalogo interfacciato da una maschera di ricerca di tipo GUI.
- Interfaccia grafica accessibile attraverso l'utilizzo di un browser (Netscape, Explorer, Opera ...)
- Protocollo HTTP
- Linguaggio HTML
- A volte la semplicità di approccio limita la possibilità di raffinare la ricerca.

# Modalità grafica - Web



# Protocollo Z39.50

- Il protocollo Z39.50, è uno standard ANSI/NISO ed ISO 23950. E' mantenuto e regolato dalla Library of Congress. E' strutturato per facilitare l'interoperabilità tra sistemi in rete, rende cioè possibile consultare le basi dati di qualsiasi biblioteca, quale sia il suo sistema di database.
- Questo protocollo, basato sull'architettura *client/server*, specifica le strutture dei dati e le regole di scambio, che permettono ad un sistema client di effettuare ricerche nelle basi dati di un sistema server e poter consultare i records risultanti. Le potenzialità di questo protocollo permettono la consultazione di basi dati molto diverse tra loro attraverso un unico linguaggio di interrogazione e con un'unica interfaccia.

# Gateway Http-Z39.50

- Le interfacce di interrogazione degli Opac Z.39.50 vengono realizzate tramite dei gateway che non richiedono l'uso di specifici Zclient ma che sono consultabili tramite qualsiasi browser web.
- In questo modo si perdono alcune delle caratteristiche tipiche del protocollo riducendo le possibilità di quella che è probabilmente la caratteristica più interessante dello Z39.50, ovvero la possibilità di «memorizzare» la sessione in corso e i suoi risultati.



# OPAC SBN – Gateway Z39.50

ICCU Home Page - Microsoft Internet Explorer

File Modifica Visualizza Preferiti Strumenti ?

Indietro Cerca Preferiti Multimedia

Indirizzo <http://opac.sbn.it/cgi-bin/IccuForm.pl?form=WebFrame> Vai Collegamenti »

**Z39.50**

ICCU  
ICCU  
ICCU

► Ricerca Base

► Libro Moderno

► Libro Antico

► Musica

Home

Alto

Indietro

Avanti

E-Mail

Accessi

**Istituto Centrale per il Catalogo Unico**

**Indice SBN**

*Modulo di Ricerca Base*

(Legenda: ● Lista delle frasi: ● Lista delle parole)

**Autore** ● ●

**Titolo** ● ●

**Soggetto** ● ●

**Classificazione:**

**Numero** ●

**Descrizione** ● ●

**Tutti i campi**

Cerca ►

Cerca ►

Ordinamento per  Max Documenti: 20 Formato Scheda: SUTRS

Ricerca Base

Microsoft PowerPoint - [C...]

ICCU Home Page - Mic...

Internet

11.43

# META-OPAC

Sono strumenti che permettono di interrogare più Opac contemporaneamente.

Vi è quindi una fusione virtuale di Opac web grazie ad una nuova interfaccia che si interpone alle altre.

I metaopac interrogano diverse basi dati, dunque non permettono ricerche particolarmente raffinate e mostrano i dati in formati diversi.

Un esempio è il Metaopac Azalai Italiano (MAI)


# MetaOpac Azalai

MAI. Ricerca globale - Microsoft Internet Explorer



File Modifica Visualizza Preferiti Strumenti ?

Indietro → → → Cerca Preferiti Multimedia

Indirizzo <http://azalai.cilea.it/mai/> Vai Collegamenti »



## MAI. Ricerca globale

una collaborazione tra AIB  e CILEA 

**Interrogazione cumulativa dei cataloghi delle biblioteche italiane** [info](#)  
Cumulative query in Italian Internet library catalogues

*Immettere uno o più termini per almeno un campo di ricerca* [info](#)

TITOLO:	<input type="text"/>	<a href="#">info</a>
AUTORE:	<input type="text"/>	<a href="#">info</a>
TIPO DI DOCUMENTO:	<input type="text" value="tutti i documenti"/>	<a href="#">info</a>

solo risposte positive ☒ [info](#)



*premere una sola volta e attendere*

Questo metaOPAC è indicato per la ricerca di [documenti poco comuni non trovati in altri cataloghi italiani](#).

Attraverso questa maschera vengono interrogati direttamente tutti i [cataloghi attualmente connessi al MAI](#); sono disponibili anche una [ricerca selettiva](#) per aree geografiche, tipi di biblioteche e tipi di documenti; e una [ricerca per regione](#). Tutti i cataloghi in rete italiani, compresi quelli non interrogabili dal MAI, possono inoltre essere consultati singolarmente a partire dal [repertorio degli OPAC italiani](#).

Per segnalare problemi e malfunzionamenti ed inviarci commenti e suggerimenti vi invitiamo a scrivere a [<azalai@cilea.it>](mailto:azalai@cilea.it).

Copyright AIB-CILEA 1997, 2000. Repertorio degli OPAC italiani creato **1997-02-09**; metaOPAC MAI progettato a partire dall'estate 1998, disponibile al pubblico dal 1999-05-18; pagina creata il **2000-09-21**; ultimo aggiornamento di questa pagina **2003-11-27** a cura della [redazione di "OPAC italiani"](#)

Start   Microsoft PowerPoint - [C...] MAI. Ricerca globale - ... 150.146.3.20 Internet 11.42

# La ricerca in OPAC

In genere sono disponibili due tipi di ricerca: per liste (browse) o per parole (find).

Nella ricerca per liste viene visualizzato un elenco di termini che iniziano con la parola cercata o con parole simili che nell'ordine alfabetico si trovano vicino a quella richiesta.

Nella ricerca per parole vengono mostrati tutti i record che contengono la parola cercata.

Nella ricerca per parole è possibile ricorrere agli operatori booleani (in genere *and*, *or* e *not*, a volte anche altri) e ai caratteri jolly che permettono di raffinare la ricerca.



# Operatori logici (booleani)

- AND o + (intersezione): si usa per rintracciare i record che soddisfano entrambi i criteri richiesti.
- OR (unione): si usa per rintracciare i record che soddisfano almeno uno dei criteri richiesti.
- NOT o AND NOT (complemento): si usa per rintracciare i record che soddisfano un determinato criterio escludendo quelli che però ne soddisfano un altro.

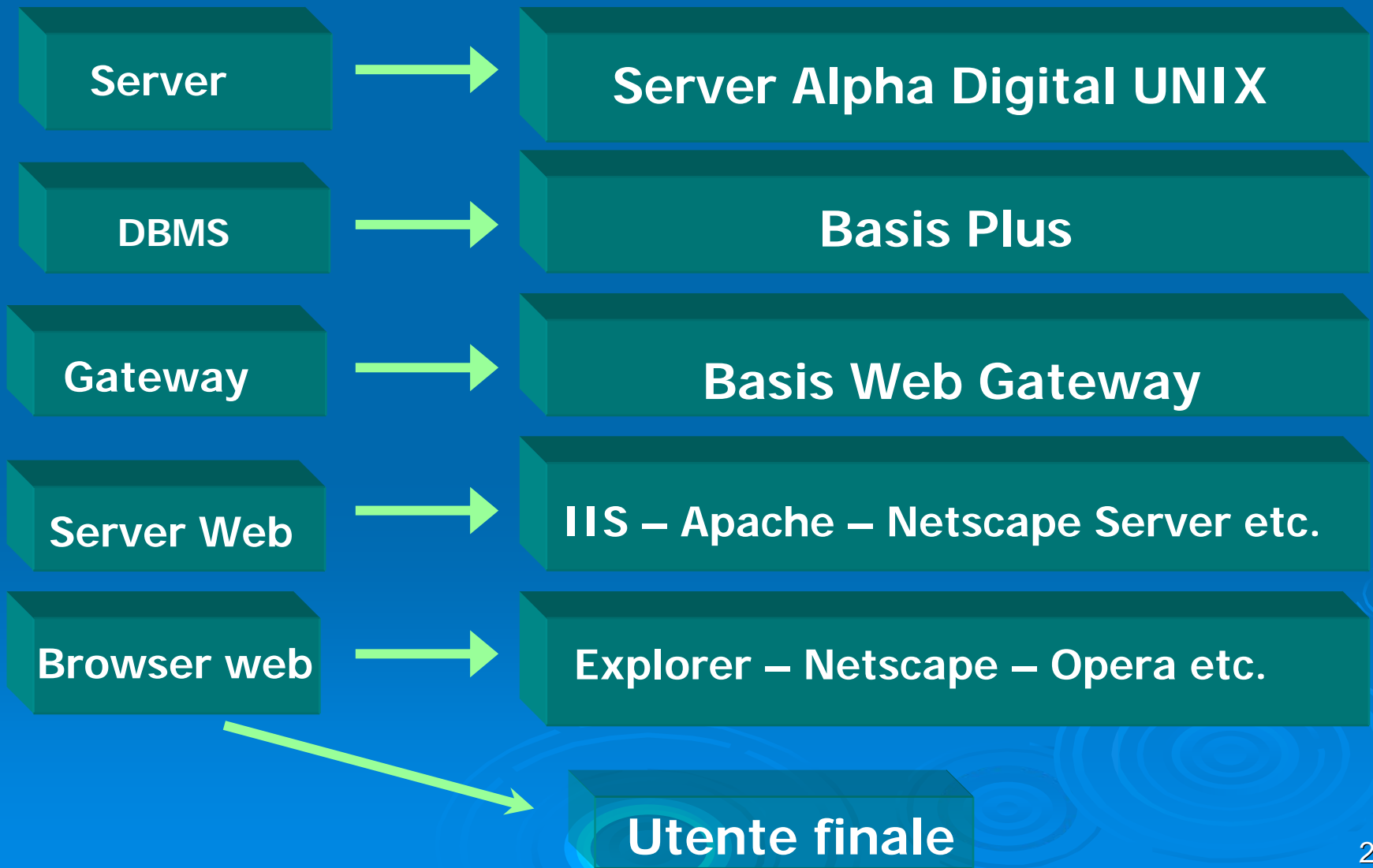
# Caratteri Jolly

I cosiddetti *caratteri jolly* sono dei simboli che possono corrispondere a qualsiasi carattere, oppure a una sequenza di caratteri di qualsiasi lunghezza. Anche nei simboli usati come caratteri jolly gli Opac variano molto, e occorre appurarli in ciascun caso leggendo le relative istruzioni.

Ecco i casi in cui si usano:

- troncamento: è un modo per evitare di dover digitare lunghe sequenze di termini alternativi aventi una stessa radice, separati dall'operatore OR (carattere jolly usato: asterisco, dollaro, due punti, o altro).
- mascheramento: si sostituisce un determinato numero di caratteri all'interno della parola (carattere jolly usato: soprattutto asterisco, o punto interrogativo).

# OPAC Bice – Architettura



# Conversione catalogo

## Pianificazione

- ◆ La decisione di convertire il catalogo preesistente in un supporto elettronico comporta una serie di scelte di vario tipo, di natura tecnica, strategica, gestionale, e finanziaria.

# Scelta del metodo

- ◆ La conversione può essere realizzata in diversi modi.
- ◆ Il sistema migliore dipende da fattori di varia natura:
  - Il tipo e la dimensione della collezione
  - Le esigenze della biblioteca
  - I fondi disponibili
  - Il tempo e il personale a disposizione

# Tecniche di conversione

- ◆ Si distinguono in linea generale due casi:
  - L'esistenza di record già disponibili su supporto elettronico
  - La creazione ex-novo di record bibliografici



# Caricamento di record disponibili

## Catalogazione derivata

- ◆ **Utilizzo di una base dati fuori linea offerta da fornitori commerciali. Ad es.**
  - La grande b.d. mondiale OCLC (Online Computer Library Center) in cui la biblioteca ricerca i record da convertire e riceve su supporto elettronico i record trovati in vari formati.
  - Utilizzo di cd-rom bibliografici nazionali e internazionali come BNI, il catalogo della BNCf e altri.
  - Infine si possono catturare anche i record da un catalogo collettivo di tipo locale in linea o residente su cd-rom.

# Caricamento di record disponibili

## Catalogazione derivata

- ◆ Il ricorso a base dati fonte comporta innanzitutto operazioni di ricerca, identificazione ed estrazione secondo un formato che sia facilmente integrabile nel proprio Opac.
- ◆ E' essenziale che il tasso di copertura dei record recuperati giustifichi l'operazione (almeno il 60% di record trovati)



# Catalogazione derivata

## Utilizzo di base dati fonte - II

- ◆ Vantaggi: costi ridotti per record, facilmente calcolabili e tempi ristretti
- ◆ Svantaggi: Lavoro di identificazione e selezione, casi di ambiguità. Necessità di interventi sui record. Vi sarà inoltre sempre una parte di record da immettere, quando non sia possibile recuperarli.
- ◆ Necessità di uso di un formato di scambio

# Cos'è un "formato"

- ◆ Si definisce "formato" la struttura logica degli oggetti digitali.
- ◆ Perché un software possa riconoscere e gestire un oggetto digitale è necessario che i dati siano strutturati secondo un formato comune che li etichetti e assegni loro un preciso significato e che consenta a tutti i software di gestione di riconoscerli e importarli in modo corretto.
- ◆ I "formati di scambio" dei dati sono dedicati allo scambio di dati fra sistemi differenti.

# UNIMARC

- ◆ **UNI**versal **MA**chine **R**eadable **C**atalogue
- ◆ È un un sistema di codifica internazionale per la leggibilità dei dati bibliografici registrati su supporto elettronico ed un formato macchina standardizzato per lo scambio di record bibliografici.

# UNIMARC

- ◆ Conforme a ISO 2709
- ◆ Conforme alle ISBD
- ◆ Compatibile con le RICA
- ◆ Molto analitico e flessibile
- ◆ Adottato da BNI e da SBN



## 2 - Creazione di record bibliografici

- ◆ Può essere realizzata in diversi modi:
  - Immissione manuale delle schede in una base dati
  - Riduzione a immagine del catalogo
  - Lettura ottica e riconoscimento più o meno automatico dei caratteri e dei campi bibliografici

# Immissione manuale delle schede

- ◆ Si tratta di un sistema sempre meno utilizzato, da escludere per volumi consistenti di record.
- ◆ Vantaggi: i costi sono bassi.
- ◆ Svantaggi: il lavoro richiede tempi lunghi e spesso è accompagnato da perdita di interesse e importanza dell'intera operazione.

# Ricorso a tecnologie di lettura ottica

- ◆ Si può inquadrare in due ambiti diversi:
  - Digitalizzazione del catalogo come immagine
  - Riconoscimento ottico dei dati

# Digitalizzazione del catalogo come immagine

- ◆ La scansione ottica delle schede e quindi la memorizzazione delle immagini delle schede stesse su cd-rom pur non essendo una vera conversione su OPAC può essere una misura per la salvaguardia e la conservazione del catalogo stesso.
- ◆ Si può creare una sorta di indicizzazione delle immagini con un apposito software che consente di accedere a partire da una ricerca per vari modi di accesso alle varie schede e scorrerle sequenzialmente o a salti così come avviene per la consultazione manuale.



# Digitalizzazione del catalogo come immagine - II

The screenshot displays a web-based digital catalog interface. At the top, a banner reads "Camera dei Deputati - Biblioteca" and "Cataloghi". Navigation links include "HOME", "INFO CATALOGO", and "GUIDA ONLINE".

On the left sidebar, under "Cataloghi 1848-1983", there are search filters for "1956-1983" (Dizionario) and "1848-1955" (Autori, Titoli, Metodico (schede), Metodico (volumi)). A search box contains the text "Manzoni" and a "Ricerca elenco" button is below it. Further down, under "Catalogo 1984-", there is a "catalogo on-line" link.

The main content area features a table with the following data:

da	a	tot.schede	scheda
manzini	maran	207	76

Below the table, a large image of a handwritten manuscript is displayed. The text in the manuscript reads:

*Manzoni*  
*I promessi sposi. Con uno studio di Nicolo Bonuma*  
*100.*  
*Milano, Ist. Est. Ital, s. I, in 16°, p. 628, con rit.*  
*In: Classici italiani Martini, s. I, n. 518.*

At the bottom of the manuscript image, there is a date stamp: "Ist. Est. Ital, 3/X-1913" and a handwritten number "326/3".

Below the manuscript image, there is a "zoom" button and a set of navigation controls (back, forward, etc.). A "vai" button with a text input field is also present.

# Digitalizzazione del catalogo come immagine - III

- ◆ Vantaggi: Costo ridottissimo dell'operazione e da un eventuale risparmio di spazio in caso della rimozione del catalogo cartaceo. Importante la possibilità di accedere al catalogo-immagine in remoto via web.
- ◆ Svantaggi: Mancanza di disponibilità dei record nell'OPAC della biblioteca e quindi separazione tra le risorse informative con l'impossibilità di fare una ricerca di documenti sull'OPAC.

# Riconoscimento ottico dei dati

- ◆ E' una tecnica considerata a oggi ancora sperimentale per la difficoltà di una sua reale applicabilità ma promettente nell'ottica di un futuro sviluppo tecnologico.
- ◆ Si tratta, dopo una lettura ottica – tramite scanner – di far riconoscere i caratteri della scheda cartacea da opportuni software (OCR).

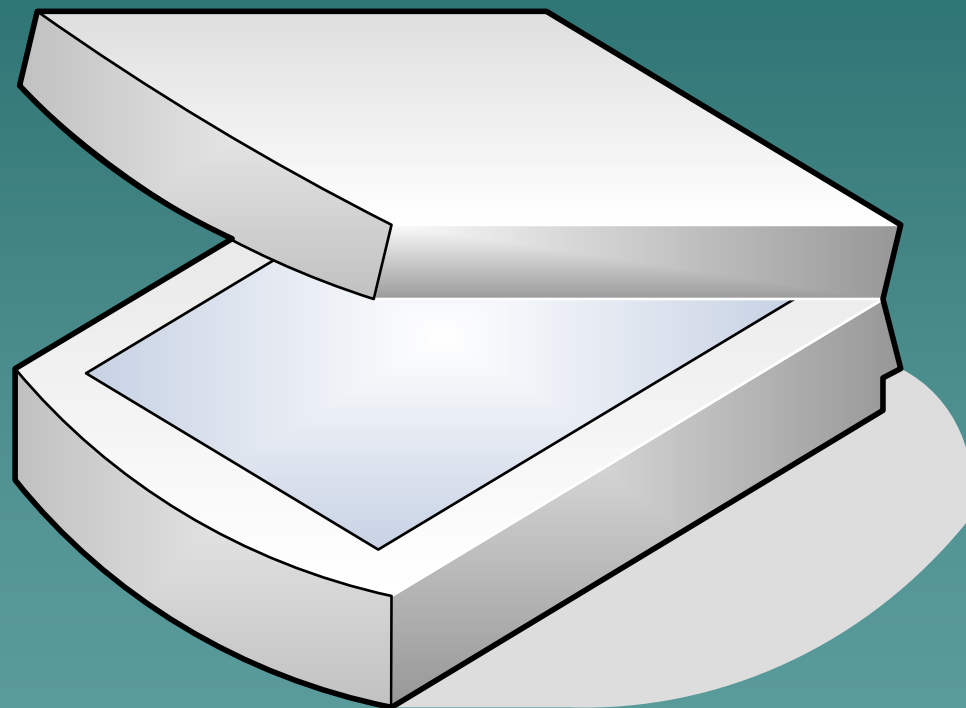
# Riconoscimento ottico dei dati II

- ◆ Il secondo passaggio è l'interpretazione dei caratteri riconosciuti dall'OCR come elementi bibliografici da inserire nei rispettivi campi della base dati.
- ◆ I software OCR si scontrano tuttora con le difficoltà dovute alla molteplicità, diversa dimensione, tipologia e dislocazione dei caratteri usati nelle schede cartacee.

# Riconoscimento ottico dei dati III

- ◆ Vantaggi: Trasformazione automatica del catalogo cartaceo in documento elettronico
- ◆ Svantaggi: Necessità di pesanti interventi di revisione e completamento del processo automatico di riconoscimento e memorizzazione dei dati.

# Scanner e OCR



# Scanner - Tipologie

## ◆ Scanner manuale:

- Di piccole dimensioni, occorre avere mano ferma e tenere velocità costante per acquisire l'originale in modo accurato. Non permettono di digitalizzare originali superiori ai 10-12 cm di larghezza.

## ◆ Scanner piano:

- Molto simile ad una piccola fotocopiatrice, si posiziona l'originale con la faccia rivolta verso il basso sulla lastra di vetro. Vanno dal formato A4 in su. Una sorgente di luce montata su un carrello mobile viene fatta passare sotto il vetro e illumina fortemente l'originale; un sensore lineare anch'esso mobile digitalizza l'immagine. Hanno una risoluzione che va dai 300 dpi in su.

# Scanner – Tipologie

- ◆ Scanner con ADF:
  - Concettualmente molto simili ai fax, la lampada e il sensore che digitalizza l'immagine sono fermi, mentre è il foglio con l'immagine da acquisire che si muove con un sistema a rulli. Questo tipo di scanner è utile soprattutto per l'acquisizioni di documenti in serie.
- ◆ Scanner per diapositive:
  - Servono esclusivamente per digitalizzare diapositive e negativi da 35 mm. Molto costosi, offrono una elevata qualità di scansione.
- ◆ Scanner a tamburo:
  - Sono utilizzati per uso professionale in tipografie, studi grafici, case editrici etc.



# Scanner - Funzionamento

Gli scanner funzionano utilizzando il principio della **riflessione**. L'originale viene posto con la faccia rivolta sul piano dello scanner e quando la testina di scansione si sposta lungo il piano, la luce generata da una apposita lampada viene diretta sull'originale.

I sensori presenti sulla testina misurano la quantità di luce riflessa che viene poi convertita dai sensori in segnali digitali direttamente trasferiti al computer. I dati ricevuti dal computer vengono passati ad un programma applicativo che li interpreta per visualizzare l'immagine dell'originale sullo schermo.

# Scanner - Risoluzione

- ◆ La risoluzione ottica dello scanner si misura in DPI (punti pollice). Dipende da:
  - Massima accuratezza ottenibile dal sensore CCD.
  - Accuratezza meccanica del motore passo-passo che guida la testina di scansione.
- ◆ Il sensore CCD:
  - E' formato da migliaia di elementi fotosensibili, il cui numero determina la risoluzione dello scanner.
  - Es: se lo scanner è largo 8,5 pollici (21,59 cm.) e il ccd contiene 2540 elementi la massima risoluzione è  $2540/8,5 = 300 \text{ dpi}$

# OCR – Principi tecnici

- ◆ In linea generale un metodo OCR tradizionale prevede una fase di acquisizione, tramite una opportuna ottica, di una immagine testo che viene digitalizzata e rappresentata da una "bitmap" appunto digitale.
- ◆ All'interno della bitmap digitale il software di elaborazione dell'immagine ha lo scopo di distinguere la zona di immagine relativa ai singoli caratteri.

# OCR Principi tecnici - II

- ◆ Una volta individuati i caratteri, o meglio la loro immagine digitale, devono essere "riconosciuti" ovvero deve essere attribuito loro il corrispondente carattere a stampa in modo da poter ricostruire il testo, ora in forma digitale.
- ◆ I punti deboli di un sistema OCR dipendono dalla irregolarità dei caratteri a stampa, che possono presentarsi con sfumature di colore o forme più o meno diversi, rendendo difficile una definizione standard del carattere. In particolare questo è vero per testi a stampa antichi o in precarie condizioni di conservazione.

**Grazie  
dell'attenzione**